DOCUMENT RESUME

ED 079 409 TM 003 013

AUTHOR	Boldt, Robert F.
TITLE	Factor Analysis of Law School Grades.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-73-42
PUB DATE	Jun 73
NOTE	24p.; Draft Copy
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Bias; *Factor Analysis; *Grades (Scholastic); Grading; Higher Education; *Law Instruction; Scores; Statistical Analysis; *Teacher Influence; Technical Reports; Test Results

ABSTRACT

Grades of 116 law school students were obtained from transcripts, along with LSAT, Writing Ability and Background scores. The grades were factored, and an analysis of the goodness of fit of one- through four-factor systems indicated that the system of grades was essentially one factor in nature. A plot of factor loadings for the various courses made for the two- and three-factor systems indicated that a few courses were somewhat different than the rest, but the reason was not clearly identifiable. The two-factor system was decomposed into two uncorrelated components, one which correlated .96 with the single factor scores and the other which correlated zero with the single factor scores. The latter component had a multiple correlation of .03 with LSAT, Writing Ability and Background scores. These analyses produced no persuasive evidence for more than one factor. The results agree with those of a similar study on the business school context. The residuals produced were used to study the possible existence of instructor bias. Two instructors were found with residuals significantly different from zero. One was found to be a slightly easy grader, and the other was found to be a hard grader. (Author)

RESEARCH BULLETIN

RB-73-42

FACTOR ANALYSIS OF LAW SCHOOL GRADES

Robert F. Boldt

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
June 1973

FACTOR ANALYSIS OF LAW SCHOOL GRADES

Robert F. Boldt

Abstract

Grades of 116 law school students were obtained from transcripts, along with LSAT, Writing Ability, and Background scores. The grades were factored, and an analysis of the goodness of fit of one- through four-factor systems indicated that the system of grades was essentially one factor in nature. A plot of factor loadings for the various courses made for the two- and three-factor systems indicated that a few courses were somewhat different than the rest but the reason was not clearly identifiable. The two-factor system was decomposed into two uncorrelated components, one which correlated .96 with the single factor scores, and the other which correlated zero with the single factors scores. This latter component had a multiple correlation of .03 with LSAT, Writing Ability, and Background scores. These analyses produced no persuasive evidence for more than one factor. The results agree with those of a similar study on the business school context.

The residuals produced were used to study the possible existence of instructor bias. Two instructors were found with residuals significantly different from zero. One was found to be a slightly easy grader, and the other was found to be a hard grader.

FACTOR ANALYSIS OF LAW SCHOOL GRADES

Robert F. Boldt

Educational Testing Service

Many educational researchers have concluded that the limits of validity have been reached due to the complex nature of the grade point criterion. This presumed complexity arises in part from the diverse nature of the courses that may be taken by the given student and the corresponding abilities that may be required. For example, in business education one might expect a course on the international aspects of business to require different abilities for successful performance than would a course in computer operations or controllership. In fact, one might expect to find a group of courses which could reasonably be called the "verbal group" based both on their content and their relationship between grades in these courses and scores on a test of verbal ability. Similarly one might find "technical groups" which relate to quantitative ability. At least such were the expectations when a factor analysis of graduate business school grades was conducted. At two graduate schools of business the grades of a graduating class were examined to determine the number of abilities required to account for the academic performances. At both schools, only one factor, i.e., one ability, seemed to underlie the business school grades. This was in contrast to the strong presumption that some courses would be primarily dependent on verbal abilities and other courses would depend primarily on quantitative or technical skills. It had been thought at the outset of these studies that the courses could be grouped so that courses in a single group would all require the same

balance of abilities. If more than one such group was found, then the relation between performance in courses within a group and test scores could lead to an identification of the nature of the predictable parts of the grade point criterion. If several abilities had been discovered, then the grade point average could be fractionated into relatively pure sub-criteria and test development efforts initiated to improve prediction for those courses not currently well predicted. With the single ability result, however, there was no possibility of such fractionation, and, in fact, the value of the single grade point average as a criterion was supported.

Within the context of legal education there is reason to suspect that perhaps the unidimensional character of the factor structure of grades might not obtain. There has been a changing concern in recent years with the sociological relationship between law and society, and this concern, among others, is manifested in electives or in the second and third years. It is thought that possibly the traditional concerns reflected in the first year of law training may as a group require different abilities than the broader exposures occurring later on. To the extent that this is true, different legal courses may require different balances of common abilities for excellence of performance and the grade point criterion might be modified as was suggested in the business school context. If only a single factor is found, then one may conclude that though the subject matter learned is different in content, the abilities required for its mastery are similar or are unique for each course, i.e., the change in subject matter does not imply a change in the nature of the predictors required.

Finally, there is a concern with instructor differences in grading. It is reasonable to suspect that different instructors grade with different

reliability. Teaching approaches may differ as may testing practices. For this reason, an analysis of instructor differences was also conducted in this study.

Methodological Orientation

The sequence of events in conducting the study begins with an attempt to determine the structure of ability requirements for the various courses. This is achieved by associating certain parameters with students and certain other parameters with courses and by determining how well suitable combinations of these parameters summarize course grades. For example, when three abilities are postulated, the parameters associated with the Business School course "Organization and Leadership of Enterprises" (Boldt, 1970) were found to be

$$a_1 = .007 \ , \quad a_2 = -.041 \ , \quad a_3 = -.149 \ , \quad b = 2.627$$

and the parameters associated with a particular student were

$$z_1 = 64.227 \ , \quad z_2 = 1.096 \ , \quad \text{and} \quad z_3 = .711 \ .$$

The a's, b's, and z's are combined to get an approximate grade as follows:

$$a_1z_1 + a_2z_2 + a_3z_3 + b$$

$$= (.007)\ (64.227) + (-.041)\ (1.096) + (-.149)\ (.711) + 2.627$$

$$= 2.928 \ .$$

The student whose parameters are given above actually obtained a grade of 3 (C) in "Organization and Leadership of Enterprises"; thus the three-ability system approximated his course performance quite well, leaving a residual error of only .072 (3.000 - 2.928) on the grade point scale. It

is these squared residuals, summed over all people and over all the courses they took, that the analysis minimizes.

Notice that the scores associated with people, often referred to in the factor analytic literature as factor scores, are here called ability scores. This is considered reasonable in the sense that variations in z's produce varying approximated performances in the same course. The number of abilities required by a system of course performances is inferred by finding the minimum number of postulated abilities which yield a good approximation of the grades. For example, if two abilities had been postulated in the preceding example, there would be only two a's and two z's, and the residuals would be different than when the three ability system is used to approximate the grades. The sum of squares of these residuals for the two-ability system can be compared with the sum of squares of residuals for the three-ability system to help decide if the addition of a third ability provided a better fit to the data. Of course, the three-abilit, system uses more parameters than the two-ability system; it uses as many more z's as there are students and as many more a's as there are courses. Use of so many more parameters ensures that the three-ability system will fit the data better than will the two-ability system but trivial reductions in the sum of squares of residuals will not be considered evidence which confirms the necessity to accept the third factor.

In summary of the methodological points above, there are four kinds of quantities that result from the factor analysis. One is a set of parameters, b's in the discussion above, which adjust for the difficulty of the courses. Another set of parameters, z's in the discussion above, represent the student's scores on the inferred variables which underlie the

grades. These z 's combine into weighted sums to approximate the grades, and the weights in these sums are the a 's of the discussion above; the larger a weight the greater the dependence of the grades on that ability. Finally, the goodness of fit of a particular number of factors is indicated by the sum of squares of residuals, a smaller value indicating a better fit.

Data

Data which were obtained from an American law school consisted of transcripts of law school performance for 116 students. Along with a student's transcript was included his LSAT score as well as his Writing Ability and Background scores. These data were sorted by course and 62 courses were identified for which grades for seven or more students were available. In addition, information was provided as to which instructors taught the various courses. In all, 3315 grades were the main subject of the analysis.

Analysis of Factors by Course

The first part of the analysis was intended to determine the number of factors. Table 1 presents the sums of squares of residuals for various numbers of factors and a correlation interpretation of these data. The bench-mark value taken for this analysis is the fit of a single mean value for all of the observations. The sum of squares of differences between that value and the observed grades is given in the first line of Table 1. Line two of the table gives the sum of squares of residuals when the average grade for a course is used to approximate the grades in the course; 62 such averages were calculated, there being 62 courses, and it can be noted that the sum of squares of residuals in line two is not very different than

that in line one. The difference between the two sums of squares is given in the third column of Table 1, and that difference is interpreted as a multiple correlation in the fourth column of that table. The multiple correlation interpretation is developed just as in other research which uses correlation ideas, where the squared multiple correlation is equal to the percent of sums of squares accounted for by the predictors and the sums of squares are taken around a single criterion average. As far as predicting the grades is concerned, one could replace a predictor which correlated .36 with the grades with the course average and predict just as well. Note that fitting one factor improves the fit markedly, increasing the correlation by 30 points, but after that the improvement in fit is not large. Usually, when one is selecting predictors in a regression situation, one requires that the addition of a test to a prediction set adds at least .005 to the multiple correlation coefficient for the selection of variables to continue. This increase comes about because of the regression weight for the added variable, but in the present situation one adds not onl regression weight for each course but also a factor score for each student. Each factor requires the addition of 150 or more parameters yet the gain noted after one factor in Table 1 is .1 or less.

Insert Table 1 about here

A plot of the two-factor solution a 's was developed and presented in Figure 1 to suggest substantive hypotheses even though the analysis of residuals does not establish the need for acceptance of the second factor.

Insert Figure 1 about here

Virtually all the points lie in the shaded area on the plot with the exception of the extreme deviants whose course titles are given along with the number of students enrolled and the time the course is offered. Note tnat those courses which appear deviant are not entirely from a single year, nor do they seem restricted to courses which reflect a common content theme.

In the previous paragraph the plot of the a 's for the two-factor solution was examined merely for the sake of finding interesting possible interpretations of the data. The case for examining the three-factor solutions is somewhat stronger, as can be seen by examining the column labeled "Sums of Squares Attributable to Parameters" in Table 1. Notice that in this column the gain by fitting the first factor is quite large (247.78 - 74.52 = 173.26), the gains by fitting the second factor (319.74 - 247.78 = 71.96) and the third (391.54 - 319.74 = 71.80) are about the same, and the gain by fitting the fourth factor is smaller (413.24 - 391.54 = 21.8). Some might take the equality of the reductions by the second and third factors, followed by the rather small reduction supplied by the fourth as evidence in favor of accepting the third factor. For this reason, Figure 2 was developed for interpretative purposes. Figure 2 is a plot of modified a 's obtained with the three-factor solution. The modification used is consistent with the method of extended vectors (Thurstone, 1947) which allows the examination of a three-dimensional configuration using a two-dimensional space (piece of paper). In this application of the method of extended vectors the three-dimensional plot of a 's was projected onto a plane defined by a coordinate of unity for the axis corresponding with the third factor. If the a 's describe a single factor except for a few

outliers, the plot of Figure 2 would appear essentially as a streak of points on a line with a few points off the line. It does look a little that way, but three axes have been drawn in to indicate what one might wish to interpret as three streaks, though the evidence for these three streaks seems not very strong. The points defining the extremes of the streaks have been numbered along with some outliers so that these particular points and streaks can be discussed and presented in Table 2. The courses are arranged by the vector with which they are associated, if they are not simply a deviant, and one may note that Course 950-11 is included twice as it gives an end point of both vectors A and B. The footnoted entries in Table 2 are ones which occur because the numerical operations involved in the extended vector computations require a division by a number which was very small in the case of these courses. The footnotes dictate caution in placing confidence on interpretations based on the footnoted loadings. Table 2 also contains the year in which the courses are normally taken.

Insert Figure 2 and Table 2 about here

The author's interpretation of Figure 2 and Table 2 is that the separation of vectors A and B is somewhat forced, but that perhaps vector C indicates some trend. If so, that trend seems not to be one of differentiating socially oriented, third-year courses from the more traditional first-year courses. In fact, if the deviants of Table 2 are attributed to sampling instability due to division by a small number and if vectors A and B are considered to be essentially the same, then the space would be considered

a two-dimensional space. But the two-factor solution has been examined and not found to be enlightening. The evidence for a second or third factor still does not seem convincing though interpretations of this kind are subjective and the reader is not bound by technical considerations to agree.

Table 3 is a table of intercorrelations of LSAT, Writing Ability, Background, F11 (the factor scores from the single factor analysis), and F12 and F22 (the first and second factors from the two-factor analysis). The validity of the three ... ts for predicting F11 was .39 with standard score regression weights of .36, .14, and -.13 for LSAT, Writing Ability, and Background, respectively. To examine the correlation of the predictors with the two-factor system the F12 and F22 scores were combined into a composite that was uncorrelated with F11. This composite would represent that part of the factor variance that is unrelated to the first factor--essentially new variance introduced by the inclusion of the second factor. One would be quite interested to observe a substantial relationship between this composite and the predictors because it would indicate the existence of systematic variance in the two-factor system and a need for different combinations of the predictors for the two factors. However, the multiple correlation between this composite and the predictors was only .03. From the standpoint of prediction, using the LSAT, Writing Ability, and Background, there is mainly one component of the two-factor system and that component is the part that is predictable using the tests. The multiple correlation between the two factors and F11 is .96.

Insert Table 3 about here

Analysis of Residuals for Instructor Effects

One problem in analyzing for instructor effects is that it is quite difficult to separate instructor effects from course effects. For example, if an instructor taught only one course and that course was the only one he taught, then there would be absolutely no way to separate the course effects from the instructor effects. What is needed for such an analysis is a situation where an instructor teaches a number of courses and the courses are taught by a number of instructors. The data for such an analysis would require that the transcripts carry not only the course the student took, but some indication, such as an instructor code or a section code, of whom the instructor was. In the present data this condition holds, but only partially. It holds in that the sections of a course can be distinguished if they are offered in different terms, but not if offered in the same term, unless the sections in the same term are taught by the same instructor. The course identification that appears on the transcript indicates the course and grade but not the section number.

In an attempt to circumvent these difficulties it was decided to approach the problem section by section. This was possible because sections of a course were treated as equivalent for purposes of the factor analysis; and if more than one instructor gave a course, constant differences in their grading practices could be studied by examining the residuals for students in each instructor's section. Therefore, a study of the residuals by section would, when correlated with the identity of the instructors teaching the section, identify deviant grading practices. The examination of these residuals was restricted to significance testing of the

average residual by section; the variance used for this testing was the sum of squares of residuals for the one factor system divided by 3315, the number of observations. The tests were done as if the variances were known and the t -tables were used, taking the degrees of freedom as equal to the number of observations. The formula for t was that for a test of significance from zero where the standard deviation of the observations is known. It should be pointed out that these tests are probably rather low in power since the fitting of the factor model forces the residuals to average at zero for each course. Thus, if a result is indicated as significant by the test used, it is probably a real one.

Insert Table 4 about here

In all, 153 t -tests were made with 12 being significant at the 2% level or more. Table 4 presents the comparison of the number of tests found significant for various confidence levels and the number of significant tests expected on a chance level. It can be noted that at the 2% level or less there are more significant results than might be expected by chance. Table 5 presents the course name, number, the instructor number(s), the significance level, and the average residual expressed on the grade-point scale, both as a four-point scale and on the grade-point scale at this law school, which ranged from 0 to 100. Note that in Table 5, Instructor 4 and Instructor 7 appear twice as the only instructors of a section. Instructor 4 appears to be an easy grader, and

---

Insert Table 5 about here

Instructor 7 appears to be a hard grader. Table 6 contains the course title, number of cases, and average residuals on the two scales for Instructor 4 and also for Instructor 7. The average residual for Instructor 4 is -8.7 on the law school's scale indicating that he tended to give grades almost 9 points below that predicted by the one-factor model. Instructor 7 gave grades 1.2 points above those predicted by the one-factor system. Note that his average residuals are mixed in sign and that his bias of 1.2 points is not really very large when compared with the 100-point scale.

Insert Table 6 about here

Summary

Grades of 116 law school students were obtained from transcripts, along with LSAT, Writing Ability, and Background scores. The grades were factored, and an analysis of the goodness of fit of one- through four-factor systems indicated that the system of grades was essentially one-factor in nature. A plot of factor loadings for the various courses was made for the two- and three-factor systems, and it was found that there were a few courses somewhat different than the rest but their nature was not clearly identifiable. The two-factor system was decomposed into two uncorrelated components, one which correlated .96 with the single factor scores, and the other which correlated zero with the single factor scores. This latter component had a multiple correlation of .03 with LSAT, Writing Ability, and Background scores. These analyses indicated that no persuasive evidence of more than one factor could be adduced.

The factor system did not, of course, provide an exact fit to the grades, and the residuals produced were used to study the possible existence of instructor bias. Two instructors were found whose residuals were on two occasions significantly different from zero, and the average residuals were presented for all of the sections for which these men could be identified as the instructors. One was found to be a slightly easy grader, and the other was found to be a hard grader.

The results are essentially in agreement with another study in the graduate business school context (Boldt, 1970), where data from two schools were examined and a single factor was found at each location.

References

Boldt, R. F. Factor analysis of business school grades. Research Bulletin 70-49. Princeton, N. J.: Educational Testing Service, 1970.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Table 1

Analysis of Sums of Squares of Residuals

Parameter Fit	Sums of Squares of Residuals	Sums of Squares Attributable to Parameters	R[a]
Grand Mean	577.36		
b's	502.84	74.52	.36
b's and 1 Factor	329.58	247.78	.66
b's and 2 Factors	257.62	319.74	.74
b's and 3 Factors	185.82	391.54	.82
b's and 4 Factors	164.02	413.34	.85

[a]Multiple correlation

Table 2

Course Numbers, Titles, Number of Students Taking, and the Year Taken for Courses Ranked on Figure 2

	Number of Students	Course Number	Title	Year Taken
Vector A				
	114	815[a]	Practice and Procedure I	1
	107	852[a]	Business Associations II	2
	85	828	Workmen's Compensation	2 or 3
	19	950-11	International Law	3
Vector B				
	111	860[a]	Taxation I	2
	107	817	Practice and Procedure II	2
	19	950-11	International Law	3
Vector C				
	25	930	Gov't Regulation of Business	3
	98	824	Legal Research Tutorial	1
	11	872	Land Use Planning	3
	101	855	Corporation Finance Law	3
Deviants				
	13	950-25[a]	Fiduciary Administration	3
	7	825[a]	Legislation	2 or 3
	52	885[a]	Creditors' Rights	2
	111	876[a]	Trusts and Estates	1

[a]Unstable due to small loadings on Factor 3.

Table 3

Correlations of Predictor and Factor Scores

	LSAT	Writing	Background	F11[a]	F12[a]	F22[a]
LSAT	1.00	.56	.54	.36	.10	-.33
Writing		1.00	.54	.27	.08	-.26
Background			1.00	.14	.02	-.10
F11				1.00	.35	-.91
F12					1.00	-.05
F22						1.00

[a]F11, F12, and F22 are defined in the text.

Table 4

Table of Confidence Levels of *t*-Tests

	$p > .2$	$.2 > p > .1$	$.1 > p > .05$	$.05 > p > .02$	$.02 > p$
Observed	129.0	7.0[a]	3.0	2.0	12.0
Expected	122.4	15.3	7.7	4.6	3.0

[a]Number of significance tests.

Table 5

Data on Sections with Significant t-Tests of Average Residuals

Course Number	Title	Instructor Number	Significance Level	Number Cases	Average Residual: 4-Point Scale	Average Residual: Law School's
815	Practice & Procedure	25, 7	.01	3	-1.35	-33.7
815	" "	16, 31	.01	8	.48	12.0
828	Workman's Comp.	10	.01	14	.30	7.6
880-2	Commercial Law	18, 12	.02	37	-.14	-3.5
961-1	Moot Court	6	.01	17	.27	6.9
961-1	" "	U[a]	.01	4	-.78	-19.5
961-2	Legal Aid	7	.01	10	.44	11.1
961-2	" "	7	.02	14	.22	5.6
961-3	Legal Aid Intern	4	.001	9	-.59	-14.7
961-3	" " "	36	.01	6	.57	14.2
961-3	" " "	4	<.001	9	-.57	-14.2
961-6	Research Problems	U[a]	.02	14	.22	5.6

[a]Unknown.

Table 6

Course Data for Instructors 4 and 7

Course Title	Number Cases	Average Residual 4-Point Scale	Average Residual Law School's Scale
Instructor 4			
Legal Aid Intern	9	-.59	-14.7
" " "	9	-.57	-14.2
" " "	7	-.03	-.6
Tax Planning	1	-.29	-7.2
Legal Aid	3	.28	7.1
Average		-.35	-8.7
Instructor 7			
Federal Courts & Federal Systems	35	0	0
Relational Torts	16	.05	1.3
" "	3	-.03	-.7
" "	39	-.03	-.8
Legal Aid	10	.44	11.1
" "	14	.22	5.6
" "	7	-.29	-7.2
Law Journal Note Editing	6	.08	1.9
" " " "	1	-.07	-1.8
Practice Court	1	.26	6.6
Average		.05	1.2

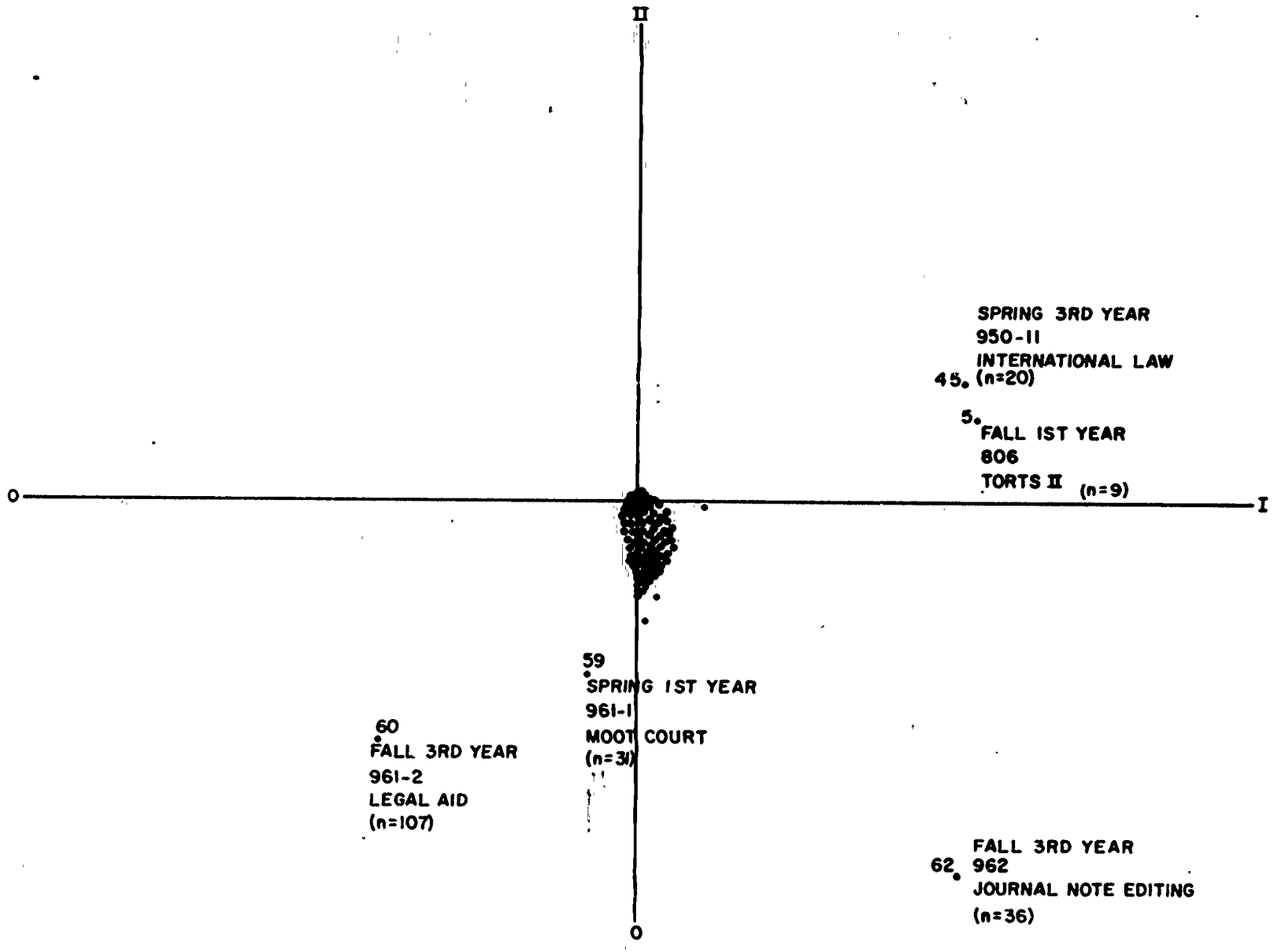
II
0
I
0
SPRING 3RD YEAR
950-11
INTERNATIONAL LAW
45. (n=20)
5.
FALL IST YEAR
806
TORTS II (n=9)
59
SPRING IST YEAR
961-1
MOOT COURT
(n=31)
60
FALL 3RD YEAR
961-2
LEGAL AID
(n=107)
FALL 3RD YEAR
62. 962
JOURNAL NOTE EDITING
(n=36)

FIGURE I

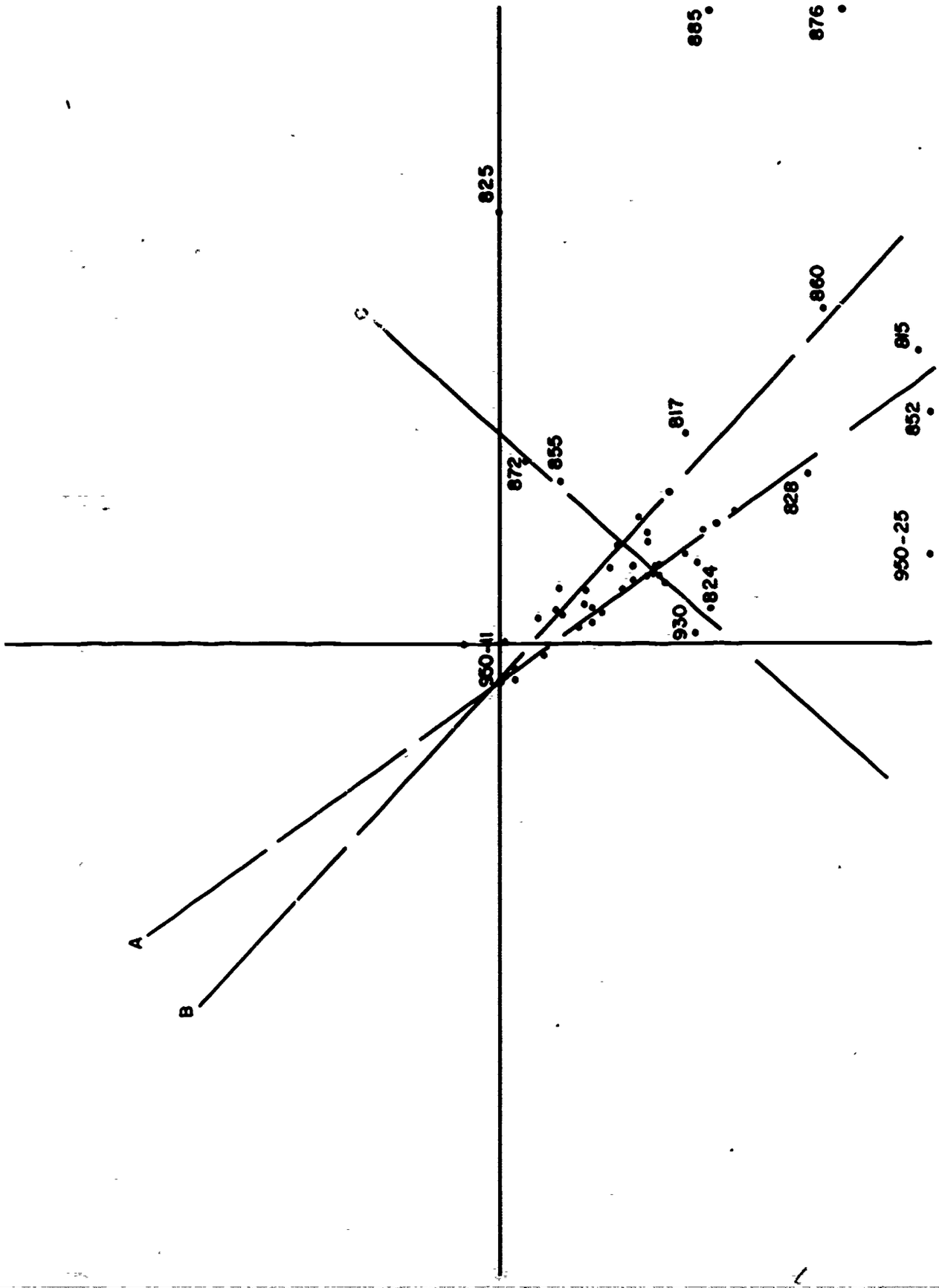

FIGURE 2